NORTH-HOLLAND

Computing the Joint Spectral Radius

Gustaf Gripenberg*
University of Helsinki
Department of Mathematics
P.O. Box 4
00014 University of Helsinki, Finland

Submitted by Richard A. Brualdi

ABSTRACT

This paper presents algorithms for finding an arbitrarily small interval that contains the joint spectral radius of a finite set of matrices. It also presents a numerical criterion for verifying in certain cases that the joint spectral radius is the maximum of the spectral radii of the given matrices. Error bounds are derived for the case where calculations are done with finite precision and the matrices are not known exactly. The algorithms are implemented and applied to estimate Hölder exponents of the orthonormal wavelets $_N\phi$ constructed by Daubechies for $3 \leqslant N \leqslant 8$.

1. INTRODUCTION

The purpose of this note is to describe and study an algorithm for calculating the joint spectral radius of a finite set of matrices. One encounters this problem in the theory of wavelets, for example, when one wants to determine the Hölder exponent of the solution of a dilation equation (see [5] and [6]). The algorithm in question extends some ideas used in [5]. We prove that in a finite (but unknown) number of steps one can find, without having to resort to brute force methods, an arbitrarily small interval that contains the spectral radius. We also give a testable criterion showing that under certain

*E-mail: Gustaf.Gripenberg@helsinki.fi.

LINEAR ALGEBRA AND ITS APPLICATIONS 234:43–60 (1996)

0024-3795/96/$15.00
SSDI 0024-3795(94)00082-O

conditions the joint spectral radius is the maximum of the spectral radii of the given matrices. Furthermore we study the problem of adapting these algorithms to cases where either the matrices are not known exactly or the calculations are done to finite precision, or both, and still get useful upper and lower bounds for the joint spectral radius.

Denote by $\mathbb{C}^{n \times n}$ the set of all $n \times n$ (complex) matrices, and let $\|\cdot\|$ be some matrix norm. If $M \in \mathbb{C}^{n \times n}$, then $\rho(M) \overset{\text{def}}{=} \sup\{|\lambda| \mid \lambda \in \sigma(M)\}$ is the spectral radius of M, where $\sigma(M)$ is the spectrum, i.e., the set of eigenvalues, of M. Let S be some bounded subset of $\mathbb{C}^{n \times n}$. Then for each $m \geqslant 1$ we let

$$\hat{\rho}_m(S) \overset{\text{def}}{=} \sup\left\{\left\|\prod_{i=1}^{m} M_i\right\|^{1/m} \middle| M_i \in S,\ 1 \leqslant i \leqslant m\right\},$$

$$\check{\rho}_m(S) \overset{\text{def}}{=} \sup\left\{\rho\left(\prod_{i=1}^{m} M_i\right)^{1/m} \middle| M_i \in S,\ 1 \leqslant i \leqslant m\right\},$$

and denote by

$$\rho(S) \overset{\text{def}}{=} \limsup_{m \to \infty} \hat{\rho}_m(S)$$

the joint spectral radius of the set S. It is clear that $\rho(S)$ does not depend on the norm one has chosen; however, $\hat{\rho}_m(S)$ does depend on the norm.

2. THEORETICAL RESULTS

The following result is a combination of [1, Theorem IV] and [4, Lemma 3.1].

PROPOSITION 1. *Let S be a bounded set in $\mathbb{C}^{n \times n}$ where $n \geqslant 2$. Then*

$$\sup_{m \geqslant 1} \check{\rho}_m(S) = \limsup_{m \to \infty} \check{\rho}_m(S) = \rho(S) = \lim_{m \to \infty} \hat{\rho}_m(S) = \inf_{m \geqslant 1} \hat{\rho}_m(S).$$

It is clear that if S is a finite set, this result gives immediately a brute force method for finding the joint spectral radius $\rho(S)$. This result gives no estimate of the rates of convergence of $\sup_{1 \leqslant m \leqslant n} \check{\rho}_m(S)$ and $\inf_{1 \leqslant m \leqslant n} \hat{\rho}_m(S)$ to $\rho(S)$ as $n \to \infty$, and it is an open problem to establish such convergence rates. In practice this convergence can be quite slow; see the example in [6]. Note also that it is possible that $\liminf_{m \to \infty} \check{\rho}_m(S) < \rho(S)$. To see this, take

$$S = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\},$$

in which case it is easy to see that $\check{\rho}_{2k}(S) = 1$ but $\check{\rho}_{2k-1}(S) = 0$ for all $k \geqslant 1$.

There is also a conjecture, raised in [7], that there is always some finite m with $\check{\rho}_m(S) = \rho(S)$. However, [7] presents examples for every $m > 1$ of finite sets S of 2×2 matrices with $\check{\rho}_k(S) < \rho(S)$ for $1 \leqslant k \leqslant m - 1$ but $\check{\rho}_m(S) = \rho(S)$.

There are a number of situations where one can directly find the joint spectral radius as $\check{\rho}_1(S) = \sup_{M \in S} \rho(M)$ or reduce the calculation to a simpler one (see [1, Lemma II] and [6, Lemma 2]):

PROPOSITION 2. *Let S be a bounded set in $\mathbb{C}^{n \times n}$ where $n \geqslant 2$, and let $V \in \mathbb{C}^{n \times n}$ be invertible. Then*

$$\rho(S) = \rho(V^{-1}SV). \tag{1}$$

Furthermore, if every $M \in S$ is upper triangular, or if every $M \in S$ is Hermitian, then $\rho(S) = \check{\rho}_1(S) = \sup_{M \in S} \rho(M)$. If every $M \in S$ is upper block triangular with square block matrices $B_1(M), \dots, B_k(M)$, with sizes independent of M, on the diagonal, then $\rho(S) = \max_{1 \leqslant j \leqslant k} \rho(S_j)$, where $S_j = \{B_j(M) | M \in S\}$.

Observe that even if one cannot reduce all elements in S simultaneously to upper triangular or Hermitian form, one can still invoke (1) in order to reduce the norms of the matrices in S in the hope that this will speed up the algorithm to be described below.

If one wants to find an approximation of the joint spectral radius without using a brute force method, one approach is to use a branch-and-bound method that disregards all products known not to be the ones determining the supremum in the definition of $\hat{\rho}_m(S)$, an idea mentioned in [5]. A precise version of this branch-and-bound approach is given below, where we use the notation

$$\Pi(X) = \prod_{i=1}^{m} M_i \quad \text{and} \quad p(X) = \min_{1 \leqslant j \leqslant m} \left\| \prod_{i=1}^{j} M_i \right\|^{1/j},$$

where $X = (M_1, M_2, \dots, M_m) \in S^m$ and $S \subset \mathbb{C}^{n \times n}$.

THEOREM 3. *Assume that S is a finite subset of $\mathbb{C}^{n \times n}$ where $n \geqslant 2$ and let $\delta \geqslant 0$. Take*

$$T_1 = S, \qquad \alpha_1 = \max_{M \in S} \rho(M), \qquad \beta_1 = \max_{M \in S} \|M\|,$$

and define recursively for $m \geqslant 2$

$$T_m = \{(X, M) \in S^m | X \in T_{m-1}, M \in S, p((X, M)) > \alpha_{m-1} + \delta\},$$

$$\alpha_m = \max\left\{\alpha_{m-1}, \sup_{Y \in T_m} \rho(\Pi(Y))^{1/m}\right\},$$

$$\beta_m = \min\left\{\beta_{m-1}, \max\left\{\alpha_m + \delta, \sup_{Y \in T_m} p(Y)\right\}\right\}.$$

Then $\alpha_m \leqslant \rho(S) \leqslant \beta_m$ for each $m \geqslant 1$ and $\lim_{m \to \infty}(\beta_m - \alpha_m) \leqslant \delta$.

The convergence rate of Theorem 3 depends on the choice of norm, and this choice should depend on S. In general it is an open problem how to choose a "good" norm, or to obtain a bound on the convergence rate. Below we discuss a useful family of norms for computer implementations, called absolute norms. This class of norms is further enlarged by invoking $\rho(S) = \rho(V^{-1}SV)$, i.e., we can study any norm obtained from an absolute norm by a similarity transformation.

Note that if one decides *a priori* that one only wants to determine the value of the joint spectral radius within an interval of length δ', then one should choose $\delta \in [0, \delta')$, but it is not obvious what the optimal choice of δ would be.

If one can calculate the matrix products exactly (and the matrices in S have rational elements), then it is straightforward to take into account the precision with which one can calculate norms, eigenvalues, and roots in the final estimate of the joint spectral radius.

We now present a criterion guaranteeing that $\rho(S) = \check{\rho}_1(S) = \max_{M \in S} \rho(M)$, which applies in particular when the matrices in S are approximately Hermitian. We say that the matrix norm $\|\cdot\|$ is *absolute* if $\|\mathrm{diag}(\lambda_1, \dots, \lambda_n)\| = \max_{1 \leqslant i \leqslant n} |\lambda_i|$. Such norms have also been called *axis-oriented*.

THEOREM 4. *Let S be a finite subset of $\mathbb{C}^{n \times n}$, where $n \geqslant 2$ and the norm $\|\cdot\|$ in $\mathbb{C}^{n \times n}$ is absolute, and let $r \overset{\text{def}}{=} \check{\rho}_1(S) = \max_{M \in S} \rho(M) > 0$. Assume that for each $M \in S$, there are an invertible matrix V and integers $q \geqslant p \geqslant 1$ (all depending on M) such that*

(i) $V^{-1}MV = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, *where* $|\lambda_1| = \rho(M)$;

(ii) *if* $(M_1, M_2, \dots, M_q) \in S^q$ *with* $M_i = M$ *for* $i = 1, \dots, k$, *where* $1 \leqslant k \leqslant p - 1$, *then there exists some number* j, $1 \leqslant j \leqslant q$, *such that*

$$\left\| \prod_{i=1}^{j} M_i \right\|^{1/j} < r; \tag{2}$$

(iii) *if* $(M_1, M_2, \dots, M_q) \in S^q$ *with* $M_1 \neq M$, *then there exists some number* j, $1 \leqslant j \leqslant q$, *such that*

$$\left(\mu_1^p \left\| Z \prod_{i=1}^{j} M_i \right\| + \mu_2^p \|V\| \, \|V^{-1}\| \left\| \prod_{i=1}^{j} M_i \right\| \right)^{1/j} < r, \tag{3}$$

where $Z = V \,\mathrm{diag}(1, 0, \dots, 0) V^{-1}$, $\mu_1 = |\lambda_1|/r$, *and* $\mu_2 = \max_{2 \leqslant i \leqslant n} |\lambda_i|/r$. *Then* $\rho(S) = \check{\rho}_1(S) = \max_{M \in S} \rho(M)$.

One cannot, of course, draw any conclusions about $\rho(S)$ if this criterion does not apply. Furthermore, in most cases one has first to perform some similarity transform so that the norms become reasonably small. If one could transform all matrices to be Hermitian, the result would follow from Proposition 2, but it is not necessary, nor always numerically possible, to achieve this exactly. Instead, what can in most cases be done when S consists of exactly two matrices is to first find a similarity transform (i.e. calculate the eigenvectors) that (at least approximately) diagonalizes one of the matrices and then find an additional similarity transform with a diagonal matrix that minimizes (approximately) the norm of the second matrix and that will leave the first, diagonalized matrix unchanged. This is, in other words, a heuristic procedure for choosing a "good" norm.

Observe that an algorithm of the type given in Theorem 3 can be used to check hypotheses (ii) and (iii). Note also that (i) can be weakened in that it is not really essential that $V^{-1}MV$ be a diagonal matrix: it would suffice for it to be a block-diagonal matrix with one block consisting of the single number λ_1 and with the norm (or at least spectral radius) of the remaining block being small enough.

If the entries of the matrix are irrational numbers or otherwise are not known exactly, or if one cannot use exact arithmetic in the evaluation of matrix products, then one has to modify the criteria of Theorems 3 and 4 so that they take into account the accumulation of errors. These issues are analyzed in detail in the next section. It is still possible that one can verify exactly the hypotheses of Theorem 4, so the only inaccuracy arises in the determination of $\max_{M \in S} \rho(M)$ in this case; see Section 4.

3. INEXACT ARITHMETIC

In this section we formulate bounds taking into account the use of inexact arithmetic. We denote by $\tilde{A} * \tilde{B}$ the floating point product of $\tilde{A}$ and $\tilde{B} \in \tilde{\mathbb{C}}^{n \times n}$, the set of complex matrices with entries that are floating point numbers. We assume that there exists a positive number ε_* such that

$$\|\tilde{A} * \tilde{B} - \tilde{A}\tilde{B}\| \leqslant \varepsilon_* \|\tilde{A}\| \, \|\tilde{B}\|, \qquad \tilde{A}, \tilde{B} \in \tilde{\mathbb{C}}^{n \times n}. \tag{4}$$

It is clear that ε_* will depend on the norm and on n; see [8, Chapter 3]. Observe that we make no assumptions about the associativity or linearity of the product $*$. To simplify the notation we write

$$\prod_{i=1}^{m}{}^{*} \tilde{M}_i \overset{\text{def}}{=} \Big(\cdots \Big(\big(\tilde{M}_1 * \tilde{M}_2 \big) * \tilde{M}_3 \Big) * \cdots \Big) * \tilde{M}_m .$$

We have the following easy result.

LEMMA 5. *Let $M_1, \dots, M_m \in \mathbb{C}^{n\times n}$ and $\tilde{M}_1, \dots, \tilde{M}_m \in \mathbb{C}^{n\times n}$ for some $n \in \mathbb{N}$, and assume that* (4) *holds. Then*

$$\left\| \prod_{i=1}^{m}{}^{*} \tilde{M}_i - \prod_{i=1}^{m} M_i \right\|$$

$$\leqslant \|\tilde{M}_1 - M_1\| \prod_{i=2}^{m} \left(\|\tilde{M}_i\| + \|\tilde{M}_i - M_i\| \right)$$

$$+ \sum_{j=1}^{m-1} \left(\left\| \prod_{i=1}^{j}{}^{*} \tilde{M}_i \right\| \left(\varepsilon_* \|\tilde{M}_{j+1}\| + \|\tilde{M}_{j+1} - M_{j+1}\| \right) \right.$$

$$\left. \times \prod_{i=j+2}^{m} \left(\|\tilde{M}_i\| + \|\tilde{M}_i - M_i\| \right) \right). \qquad (5)$$

The second problem concerns getting lower or upper bounds for the spectral radius of a matrix that one does not know exactly. In general this is a very difficult question. Here we only need reasonable estimates that can be readily used in numerical computations. For example, in applying Theorem 3 nothing is lost if one lower bound is much too small, so long as some of the estimated spectral radii are close enough to the real ones.

Thus we want to calculate the spectral radius of a matrix M (or more generally, all its eigenvalues) but we know only an approximation $\tilde{M}$ of it. We have some approximate eigenvalues of $\tilde{M}$ in the diagonal matrix $\tilde{\Lambda}$, some approximate eigenvectors in $\tilde{V}$, and an approximate inverse of $\tilde{V}$ in $\tilde{W}$. We use the following notation: $\#J$ denotes the number of elements in J, and if $J \subset \{1, 2, \dots, n\}$, then $P_J = \operatorname{diag}(\chi_J(\cdot))$ is the $n \times n$ diagonal matrix with 1 on the jth row in the diagonal if $j \in J$.

LEMMA 6. *Assume that* (4) *holds and that $\|\cdot\|$ is absolute. Let $\tilde{\Lambda} = \operatorname{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots \tilde{\lambda}_n)$ and assume that $\tilde{V}$, and $\tilde{W} \in \tilde{\mathbb{C}}^{n\times n}$ are such that*

$$\|\tilde{V} * \tilde{W} - I\| + \varepsilon_* \|\tilde{V}\| \|\tilde{W}\| < 1.$$

Suppose $J \subset \{1, 2, \dots, n\}$ *and* $J_c = \{1, 2, \dots, n\} \setminus J$. *Let* $M \in \mathbb{C}^{n \times n}$, $\tilde{M} \in \tilde{\mathbb{C}}^{n \times n}$, *and let*

$$\delta_{\tilde{W}} = \frac{\left(\varepsilon_* \|\tilde{V}\| \, \|\tilde{W}\| + \|\tilde{W} * \tilde{V} - I\|\right)\|\tilde{W}\|}{1 - \varepsilon_* \|\tilde{V}\| \, \|\tilde{W}\| - \|\tilde{V} * \tilde{W} - I\|},$$

$$\delta_M = \|M - \tilde{M}\| + \|\tilde{M} - (\tilde{V} * \tilde{\Lambda}) * \tilde{W}\|$$

$$+ \|\tilde{V} * \tilde{\Lambda}\|\left(\varepsilon_* \|\tilde{W}\| + \delta_{\tilde{W}}\right) + \varepsilon_* \|\tilde{V}\| \, \|\tilde{\Lambda}\|\left(\|\tilde{W}\| + \delta_{\tilde{W}}\right),$$

$$\delta_\lambda = \min_{i \in J, \, k \in J_c} |\tilde{\lambda}_i - \tilde{\lambda}_k|,$$

and either

$$\gamma = \delta_M\left(\|P_J \tilde{W}\| + \delta_{\tilde{W}}\right)\|\tilde{V}\|,$$

$$\eta = \delta_M\left[\|(I - P_J)\tilde{W}\| + \delta_{\tilde{W}}\right]\|\tilde{V}\|,$$

or

$$\gamma = \delta_M\left(\|\tilde{V}P_J\| + \delta_{\tilde{W}}\right)\|\tilde{W}\|,$$

$$\eta = \lambda_M\left[\|\tilde{V}(I - P_J)\| + \delta_{\tilde{W}}\right]\|\tilde{W}\|.$$

If $\delta_\lambda > (\sqrt{\gamma} + \sqrt{\eta})^2$ *and*

$$r = \frac{2\gamma\delta_\lambda}{\gamma - \eta + \delta_\lambda + \sqrt{(\gamma - \eta + \delta_\lambda)^2 - 4\gamma\delta_\lambda}},$$

then the set $\{z \in \mathbb{C} | \min_{i \in J} |z - \tilde{\lambda}_i| \leqslant r\}$ *contains* $\#J$ *eigenvalues of* M.

One could easily give more refined versions of this lemma, but it is complicated enough as it is.

Observe that if J in the result above consists of just one index $\{j\}$ and if one normalizes the 2-norm of the jth column vector in $\tilde{V}$ to be 1, then $\|P_j \tilde{V}^{-1}\|$ turn out to be what is usually called the *condition number* of the eigenvalue $\tilde{\lambda}_j$ of the matrix $\tilde{V}\tilde{\Lambda}\tilde{V}^{-1}$; see [8, Chapter 2]. Moreover, the value of the crucial number r above can be greatly affected by how one normalizes the matrix $\tilde{V}$ (that is, multiplies it on the right by a nonsingular diagonal matrix), but since any change of $\tilde{V}$ that does not have any effect when one does exact arithmetic may change the roundoff errors, it seems to be very difficult to find the optimal choice of the normalization.

One can use Lemma 6 both to find a lower bound for the spectral radius and to get upper bounds for the numbers μ_1 and μ_2 in Theorem 4(iii).

In order to implement the criterion of Theorem 4 numerically, we need one further result, quite similar to Lemma 6, that gives an upper bound for the perturbation of the eigenvectors.

LEMMA 7. *Assume that* (4) *holds and that* $\|\cdot\| = \|\cdot\|_p$, *where* $p \in \{1, 2, \infty\}$. *Let* $\tilde{\Lambda} = \operatorname{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_n)$, *and assume that* $\tilde{V}$ *and* $\tilde{W} \in \tilde{\mathbb{C}}^{n\times n}$ *are such that*

$$\|\tilde{V} * \tilde{W} - I\| + \varepsilon_* \|\tilde{V}\|\,\|\tilde{W}\| < 1.$$

Let $M \in \mathbb{C}^{n\times n}$, $\tilde{M} \in \tilde{\mathbb{C}}^{n\times n}$, *and let*

$$\delta_{\tilde{W}} = \frac{\left(\varepsilon_*\|\tilde{V}\|\,\|\tilde{W}\| + \|\tilde{W} * \tilde{V} - I\|\right)\|\tilde{W}\|}{1 - \varepsilon_*\|\tilde{V}\|\,\|\tilde{W}\| - \|\tilde{V} * \tilde{W} - I\|},$$

$$\delta_M = \|M - \tilde{M}\| + \|\tilde{M} - (\tilde{V} * \tilde{\Lambda}) * \tilde{W}\|$$
$$+ \|\tilde{V} * \tilde{\Lambda}\|\left(\varepsilon_*\|\tilde{W}\| + \delta_{\tilde{W}}\right) + \varepsilon_*\|\tilde{V}\|\,\|\tilde{\Lambda}\|\left(\|\tilde{W}\| + \delta_{\tilde{W}}\right).$$

For each $j = 1, 2, \dots, n$, *let*

$$\delta(j) = \min_{i \neq j} |\tilde{\lambda}_i - \tilde{\lambda}_j|,$$

let either

$$\gamma(j) = \delta_M\left(\|P_{\{j\}}\tilde{W}\| + \delta_{\tilde{W}}\right)\|\tilde{V}\|,$$

$$\eta(j) = \delta_M\left[\|(I - P_{\{j\}})\tilde{W}\| + \delta_{\tilde{W}}\right]\|\tilde{V}\|,$$

or

$$\gamma(j) = \delta_M\left(\|\tilde{V}P_{\{j\}}\| + \delta_{\tilde{W}}\right)\|\tilde{W}\|,$$

$$\eta(j) = \delta_M\left[\|\tilde{V}(I - P_{\{j\}})\| + \delta_{\tilde{W}}\right]\|\tilde{W}\|,$$

and assume that $\delta(j) > (\sqrt{\gamma(j)} + \sqrt{\eta(j)})^2$, *let*

$$r(j) = \frac{2\gamma(j)\delta(j)}{\gamma(j) - \eta(j) + \delta(j) + \sqrt{[\gamma(j) - \eta(j) + \delta(j)]^2 - 4\gamma(j)\delta(j)}}.$$

Suppose that $\min_{i \neq j}[|\tilde{\lambda}_i - \tilde{\lambda}_j| - r(j)] > (\|\tilde{W}\| + \delta_{\tilde{W}})\|\tilde{V}\|\delta_m$, *and let*

$$\delta_V = \frac{n^{2(p-1)/p^2}\left(\|\tilde{W}\| + \delta_{\tilde{W}}\right)\|\tilde{V}\|^2\delta_m}{\min_{i \neq j}\left[|\tilde{\lambda}_i - \tilde{\lambda}_j| - r(j)\right] - \left(\|\tilde{W}\| + \delta_{\tilde{W}}\right)\|\tilde{V}\|\delta_M}.$$

Then there exists an invertible matrix V *such that* $V^{-1}MV = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ (*i.e.*, V *is a matrix of eigenvectors of* M),

$$|\tilde{\lambda}_j - \lambda_j| \leqslant r(j), \qquad j = 1, 2, \ldots, n,$$

and

$$\|V - \tilde{V}\| \leqslant \delta_V.$$

This result is proved by using essentially the same argument as in [8, Section 2.2.4].

4. NUMERICAL APPLICATIONS

In order to compare results when using exact and inexact arithmetic, the algorithm of Theorem 3 was applied to the set

$$S = \left\{\frac{1}{5}\begin{pmatrix} 3 & 0 \\ 1 & 3 \end{pmatrix}, \frac{1}{5}\begin{pmatrix} 3 & -3 \\ 0 & -1 \end{pmatrix}\right\}, \tag{6}$$

considered by Colella and Heil in [2] and [6]. An implementation in Maple using exact arithmetic for matrix products, the 2-norm (i.e. the matrix norm subordinated to the Euclidean vector norm), and $\delta = 10^{-4}$ gave the result that

$$0.65967890 < \rho(S) < 0.65977891$$

after evaluating 1092 products, norms, and spectral radii. (To guarantee that the floating point evaluations needed for calculation norms, spectral radii, and roots gave correct answers, they where done with both 20 and 30 digit precision and it was checked that the results rounded to 10 digits were the same.) In the calculation the sets $T_1, \dots, T_{48}$ were constructed, and the largest of these contained 21 elements. (In [2] and [6] they used the l^1 norm, which explains the slower convergence.)

An implementation of Theorem 3 in Matlab, where one takes into account the finite precision by using $\varepsilon_* = 10^{-14}n^2$ in (4) and the estimate $\|M\| \leqslant (1 + 10^{-14}n^2)\operatorname{norm}(M, 2)$ (but not otherwise considering roundoff errors), gave the result that

$$0.6596789 < \rho(S) < 0.6596995$$

after evaluating 10,004 products, norms, and spectral radii. In his calculation the sets $T_1, \dots, T_{156}$ were constructed, and the largest of these contained 43 elements. After calculating 52,550 products etc. and constructing the sets $T_1, \dots, T_{243}$, one finds that

$$0.6596789 < \rho(S) < 0.6596924,$$

but after that it seems that the algorithm does not give any smaller interval (using the same accuracy as here).

When one studies the smoothness of functions satisfying a dilation equation of the form

$$\phi(\underline{x}) = 2 \sum_{k \in \mathbb{Z}} \alpha(k)\phi(2\underline{x} - k)$$

where only finitely many of the coefficients $\alpha(\underline{k})$ are nonzero, one gets the supremum of the Hölder coefficients of ϕ as the joint spectral radius of two matrices constructed (in a rather complicated way) from the sequence α; see [5, Theorem 3.1, Lemma 3.5] and [6, Theorem 5]. Applying a numerical implementation in Matlab of the criterion of Theorem 4 to the set of matrices one obtains in this way for the compactly supported wavelets ${}_N\phi$ constructed in [3], one can prove numerically that the joint spectral radius is the maximum of the spectral radii at least when $N \leqslant 8$ (in the notation of [5], $N \leqslant 15$ and odd). In this calculation we used $\varepsilon_* = 10^{-14}n^2$ in (4) and the estimate $\|M\| \leqslant (1 + 10^{-14}n^2)\operatorname{norm}(M, 2)$ (but did not otherwise consider roundoff errors, which seems to be reasonable). The conclusion one can draw is that the supremum of the Hölder exponents of ${}_N\phi$ is

N	sup
3	1.0878
4	1.6179
5	1.9690
6	2.1891
7	2.4604
8	2.7608

5. PROOFS OF THEOREMS 3 AND 4

Proof of Theorem 3. Fix $m \geqslant 1$. From the definition of $\check{\rho}_m(S)$ and Proposition 1 we immediately see that $\rho(S) \geqslant \check{\rho}_m(S) \geqslant \alpha_m$. In order to get the upper bound, we observe that it follows from the algorithm that if $X = (M_1, M_2, \dots, M_m) \in S^m$ is arbitrary, then there exists, because the sequence $\{\alpha_m\}$ is nondecreasing, a number j between 1 and m such that

$$\left\| \prod_{i=1}^{j} M_i \right\|^{1/j} \leqslant \beta_m .$$

This implies in turn that if $k > m$ and $M_i \in S$ for $i = 1, 2, \dots, k$, then there are integers $0 = j_0 < j_1 < \cdots < j_p \leqslant j_{p+1} = k$ with $j_{r+1} - j_r \leqslant m$ such that

$$\left\| \prod_{i=j_r+1}^{j_{r+1}} M_i \right\|^{1/(j_{r+1}-j_r)} \leqslant \beta_m, \qquad 0 \leqslant r \leqslant p-1.$$

But then it follows that

$$\left\| \prod_{i=1}^{k} M_i \right\|^{1/k} \leqslant \left[\prod_{r=0}^{p-1} \left(\left\| \prod_{i=j_r+1}^{j_{r+1}} M_i \right\|^{1/(j_{r+1}-j_r)} \right)^{(j_{r+1}-j_r)/k} \right] \prod_{i=j_p+1}^{k} \|M_i\|^{1/k}$$

$$\leqslant \beta_m^{j_p/k} \sup_{M \in S} \|M\|^{m/k}.$$

Since $m/k \to 0$ and $j_p/k \to 1$ as $k \to \infty$, we conclude that $\hat{\rho}(S) \leqslant \beta_m$.

It remains to show that $\lim_{m\to\infty}(\beta_m - \alpha_m) = \delta$. Suppose that this is not the case, so that there is a number $\varepsilon > 0$ such that $\beta_m \geqslant \alpha_\infty + \delta + 2\varepsilon$ for all $m \geqslant 1$, where $\alpha_\infty = \lim_{m\to\infty} \alpha_m$. By the definition of the algorithm we see that in this case we have $\hat{\rho}_m(S) \geqslant \beta_m$ for each m, and this implies that $\rho(S) \geqslant \alpha_\infty + \delta + 2\varepsilon$ as well. By Proposition 1 there is a number m and matrices $M_1, M_2, \dots, M_m \in S$ such that

$$\rho\left(\prod_{i=1}^{m} M_i\right)^{1/m} > \rho(S) - \varepsilon \geqslant \alpha_\infty + \delta + \varepsilon. \tag{7}$$

Choose the matrices M_i where $i > m$ so that $M_{i+m} = M_i$ for all $i \geqslant 1$. Next we try to find number $0 = j_0 < j_1 < j_2 < \cdots$ such that

$$\left\| \prod_{i=j_r+1}^{j_{r+1}} M_i \right\|^{1/(j_{r+1} - j_r)} \leqslant \alpha_\infty + \delta + \varepsilon, \qquad r \geqslant 0.$$

Now there are two possibilities: either we can find an infinite sequence of such numbers j_r, or there is a number j_p with $p \geqslant 0$ such that

$$\left\| \prod_{i=j_p+1}^{j_p+k} M_i \right\|^{1/k} > \alpha_\infty + \delta + \varepsilon, \qquad k \geqslant 1.$$

In the first case we conclude from the spectral radius formula that

$$\rho\left(\prod_{i=1}^{m} M_i\right)^{1/m} \leqslant \alpha_\infty + \delta + \varepsilon,$$

which is a contradiction in view of (7). In the second case we see that the sequence $(M_{j_p+1}, \dots, M_{j_p+m})$ belongs to T_m, and since $M_{i+m} = M_i$ for all i, we conclude that

$$\rho\left(\prod_{i=j_p+1}^{j_p+m} M_i\right)^{1/m} = \lim_{k\to\infty} \left\| \prod_{i=j_p+1}^{j_p+km} M_i \right\|^{1/(km)} \geqslant \alpha_\infty + \delta + \varepsilon.$$

But this means that $\alpha_m \geqslant \alpha_\infty + \delta + \varepsilon$, which is a contradiction. Thus the proof is completed. ■

Proof of Theorem 4. We apply Theorem 3 with $\delta = 0$, and we have to show that $\lim_{m\to\infty} \beta_m = r$. Assume that $(M_1, M_2, \dots, M_m) \in T_m$, let $M = M_1$, and let k be such that $M_i = M$ for $1 \leqslant i \leqslant k$, but if $k < m$ then $M_{k+1} \neq M$. We claim that if $m \geqslant p + q$, then $k > \max\{p - 1, m + q\}$ where q and p are integers depending on M. If $k \leqslant p - 1$, then we get a contradiction from (2) and the construction of the sets T_j. Assume next that $p \leqslant k \leqslant m - q$, and let j be the number in (iii) corresponding to the sequence $(M_{k+1}, \dots, M_{k+q})$. Then

$$\left\| \prod_{i=1}^{k+j} M_i \right\| \leqslant \left\| V \operatorname{diag}(\lambda_1^k, 0, \dots, 0) V^{-1} \prod_{i=k+1}^{k+j} M_i \right\|$$

$$+ \left\| V \operatorname{diag}(0, \lambda_2^k, \dots, \lambda_n^k) V^{-1} \prod_{i=k+1}^{k+j} M_i \right\|$$

$$\leqslant r^k \left(\mu_1^k \left\| Z \prod_{i=k+1}^{k+j} M_i \right\| + \|V\| \, \|V^{-1}\| \, \mu_2^k \left\| \prod_{i=k+1}^{k+j} M_i \right\| \right) < r^{k+j},$$

by (3) and the fact that $\mu_1, \mu_2 \leqslant 1$. But this contradicts the construction of T_m, and we obtain the assertion that $k > \max\{p - 1, m - q\}$. But since

$$\left\| \prod_{i=1}^{m} M_i \right\|^{1/m} \leqslant \left(\|M^k\|^{1/k} \right)^{k/m} \hat{\rho}_1(S)^{(m-k)/m},$$

and $\|M^k\|^{1/k} \to \rho(M)$ as $k \to \infty$, we get the claim of the theorem. ■

6. PROOFS OF LEMMAS 5, 6, AND 7

Proof of Lemma 5. Using (4) and standard properties of the matrix norm, we get

$$\begin{aligned}
&\left\| \prod_{i=1}^{m}{}^{*} \tilde{M}_i - \prod_{i=1}^{m} M_i \right\| \\
&\quad = \left\| \left(\prod_{i=1}^{m-1}{}^{*} \tilde{M}_i \right) * \tilde{M}_m - \left(\prod_{i=1}^{m-1}{}^{*} \tilde{M}_i \right) \tilde{M}_m \right. \\
&\qquad \left. + \left(\prod_{i=1}^{m-1}{}^{*} \tilde{M}_i \right) \left(\tilde{M}_m - M_m \right) + \left(\prod_{i=1}^{m-1}{}^{*} M_i - \prod_{i=1}^{m-1} M_i \right) M_m \right\| \\
&\quad \leqslant \left\| \prod_{i=1}^{m-1}{}^{*} \tilde{M}_i \right\| \left(\varepsilon_* \|\tilde{M}_m\| + \|M_m - \tilde{M}_m\| \right) \\
&\qquad + \left\| \prod_{i=1}^{m-1}{}^{*} M_i - \prod_{i=1}^{m-1} M_i \right\| \left(\|\tilde{M}_m\| + \|M_m - \tilde{M}_m\| \right).
\end{aligned}$$

An induction argument now gives (5). ∎

Proof of Lemma 6. We shall only consider the first definition of γ and η because the second case is completely similar.

If we let $E_R = \tilde{V}\tilde{W} - I$ and $E_L = \tilde{W}\tilde{V} - I$, then it is easy to see that

$$\|\tilde{W} - \tilde{V}^{-1}\| \leqslant \frac{\|E_L\| \, \|\tilde{W}\|}{1 - \|E_R\|}, \tag{8}$$

and it follows from (4) that we have

$$\|E_R\| \leqslant \varepsilon_* \|\tilde{V}\| \, \|\tilde{W}\| + \|\tilde{V} * \tilde{W} - I\|,$$

$$\|E_L\| \leqslant \varepsilon_* \|\tilde{V}\| \, \|\tilde{W}\| + \|\tilde{W} * \tilde{V} - I\|.$$

Thus we conclude from (8) that

$$\|\tilde{W} - \tilde{V}^{-1}\| \leqslant \delta_{\tilde{W}}. \tag{9}$$

Furthermore we have by (5) that

$$\|\tilde{V}\tilde{\Lambda}\tilde{V}^{-1} - (\tilde{V} * \tilde{\Lambda}) * \tilde{W}\| \leqslant \|\tilde{V} * \tilde{\Lambda}\|\big(\varepsilon_* \|\tilde{W}\| + \|\tilde{W} - \tilde{V}^{-1}\|\big)$$

$$+ \varepsilon_* \|\tilde{V}\| \, \|\tilde{\Lambda}\|\big(\|\tilde{W}\| + \|\tilde{W} - \tilde{V}^{-1}\|\big).$$

If we let $M_\Delta = M - \tilde{V}\tilde{\Lambda}\tilde{V}^{-1}$, we therefore get

$$\|M_\Delta\| \leqslant \|M - \tilde{M}\| + \|\tilde{M} - (\tilde{V} * \tilde{\Lambda}) * \tilde{W}\| + \|\tilde{V}\tilde{\Lambda}\tilde{V}^{-1} - (\tilde{V} * \tilde{\Lambda}) * \tilde{W}\|$$

$$\leqslant \delta_M . \tag{10}$$

Now, if $\lambda \in \mathbb{C} \setminus \{\tilde{\lambda}_1, \dots, \tilde{\lambda}_n\}$, then we have

$$\tilde{V}^{-1}M\tilde{V} - \lambda I = \tilde{\Lambda} - \lambda I + \tilde{V}^{-1}M_\Delta \tilde{V}$$

$$= (\tilde{\Lambda} - \lambda I)\Big[I + (V\tilde{\Lambda} - \lambda I)^{-1}\tilde{V}^{-1}M_\Delta V\Big],$$

and furthermore

$$\|(\tilde{\Lambda} - \lambda I)^{-1}\tilde{V}^{-1}M_\Delta V\|$$

$$\leqslant \|(\tilde{\Lambda} - \lambda I)^{-1} P_J P_J \tilde{V}^{-1}M_\Delta V\|$$

$$+ \|(\tilde{\Lambda} - \lambda I)^{-1}(I - P_J)(I - P_J)\tilde{V}^{-1}M_\Delta V\|$$

$$\leqslant \|(\tilde{\Lambda} - \lambda I)^{-1} P_J\|\big(\|P_J W\| + \|W - \tilde{V}^{-1}\|\big)\|M_\Delta\| \, \|\tilde{V}\|$$

$$+ \|(\tilde{\Lambda} - \lambda I)^{-1}(I - P_J)\|\big[\|(I - P_J)W\| + \|W - \tilde{V}^{-1}\|\big]\|M_\Delta\| \, \|\tilde{V}\|$$

$$\leqslant \frac{\gamma}{\min_{i \in J}|\lambda - \tilde{\lambda}_i|} + \frac{\eta}{\min_{i \in J_c}|\lambda - \tilde{\lambda}_i|}$$

$$\leqslant \frac{\gamma}{\min_{i \in J}|\lambda - \tilde{\lambda}_i|} + \frac{\eta}{\delta_\lambda - \min_{i \in J}|\lambda - \tilde{\lambda}_i|}.$$

Therefore it follows from a straightforward calculation that if $\min_{i \in J} |\lambda - \tilde{\lambda}_i| = r + \varepsilon$ for some sufficiently small positive number ε, then $\tilde{V}^{-1} M \tilde{V} - \lambda I$ is invertible and λ cannot be an eigenvalue of M. Since $r < \delta_\lambda$, we get the desired conclusion from a standard continuity argument; see e.g. [8]. ■

Proof of Lemma 7. It follows directly from Lemma 6 that if $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of M, then one has (after a rearrangement if necessary)

$$|\lambda_j - \tilde{\lambda}_j| \leqslant r(j), \qquad j = 1, 2, \dots, n. \tag{11}$$

Let us denote by Λ the matrix $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ and by $I + T$ the matrix of eigenvectors of $\tilde{V}^{-1} M \tilde{V}$ normalized so that T has zeros on the diagonal (if M is sufficiently close to $\tilde{V}\tilde{\Lambda}\tilde{V}^{-1}$, this is certainly true, and then one can use a continuity argument combined with the inequalities below). We can write

$$\tilde{V}^{-1} M \tilde{V} = \tilde{\Lambda} + Q, \tag{12}$$

where by (9) and by (10)

$$\|Q\| \leqslant \left(\|\tilde{W}\| + \delta_{\tilde{W}}\right) \|\tilde{V}\| \delta_M . \tag{13}$$

Multiplying both sides of (12) on the right by $I + T$, we get

$$(I + T)\Lambda = \tilde{\Lambda}(I + T) + Q(I + T).$$

If we take some indices $i \neq j$, then we get

$$T_{i,j}\left(\lambda_j - \tilde{\lambda}_i\right) = \left(Q(I + T)\right)_{i,j}.$$

Denoting by $\|\cdot\|_F$ the Frobenius (or Euclidean) norm, we deduce, because T has zeros on the diagonal, that

$$\min_{i \neq j} |\lambda_j - \tilde{\lambda}_i| \, \|T\|_q \leqslant \|Q\|_q (1 + \|T\|_q), \qquad q \in \{1, \infty, F\}. \tag{14}$$

Since $\|Q\|_F \leqslant \sqrt{n}\,\|Q\|_2$, $\|T\|_2 \leqslant \|T\|_F$, and $V - \tilde{V} = \tilde{V}T$, we get the desired conclusion from (11), (13), and (14). ■

I would like to thank the referee for several helpful suggestions to improve the presentation in this paper.

REFERENCES

1 M. A. Berger and Y. Wang, Bounded semigroups of matrices, *Linear Algebra Appl.* 166:21–27 (1992).

2 D. Colella and C. Heil, The characterization of continuous, four-coefficient scaling functions and wavelets, *IEEE Trans. Inform. Theory* 38:876–881 (1992).

3 I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* 41:909–996 (1988).

4 I. Daubechies and J. C. Lagarias, Sets of matrices all infinite products of which converge, *Linear Algebra Appl.* 161:227–263 (1992).

5 ——, Two-scale difference equations II. Local regularity, infinite products of matrices and fractals, *SIAM J. Math. Anal.* 23:1031–1079 (1992).

6 C. Heil and D. Colella, Dilation equations and the smoothness of compactly supported wavelets, in *Wavelets: Mathematics & Applications* (J. Benedetto and M. Frazier, Eds.), CRC Press, Boca Raton, Fla., to appear.

7 J. Lagarias and Y. Wang, The finiteness conjecture for the generalized spectral radius of a set of matrices, *Linear Algebra Appl.* 214:17–42 (1995).

8 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford U.P., London, 1965.

Received 29 March 1993; final manuscript accepted 19 March 1994